

# 基于贝叶斯优化和迁移学习的 CNN 算法研究

俞圳韬<sup>1</sup>, 刘万里<sup>2</sup>, 杨晓辉<sup>2</sup>, 黄玉珍<sup>2</sup>, 徐雷<sup>1</sup>, 夏吉安<sup>3</sup>

(1. 南京理工大学 计算机科学与工程学院, 江苏 南京 210094;

2. 南京市中西医结合医院, 江苏 南京 210014;

3. 南京工业职业技术大学, 江苏 南京 210023)

**摘要:**随着深度学习的快速发展,深层次的卷积神经网络在图像识别领域获得了广泛的应用,但是在实际应用中,越深的卷积神经网络在拥有强大拟合能力和优越性能的同时,也存在模型难以训练、耗费时间较长的问题。为了解决这个问题,提出使用贝叶斯优化算法对卷积神经网络的超参数进行优化,在不断的迭代过程中,得到超参数向量的最优解,然后使用最优解参数对神经网络进行训练验证,生成最优分类模型。另外,为了避免每次迭代对网络权重参数的大幅调整,网络模型架构底层使用了 VGG-16 进行迁移学习,这大大缩短了一次迭代网络收敛的时间。实验在数据集 MNIST 和 CIFAR-10 上展开,根据实验结果可知,在不牺牲图像分类模型准确率的前提下,使用贝叶斯优化算法只需要进行少数几次迭代就可以搜索到神经网络超参数的最优解,大大节省了生成最优分类模型的时间和 GPU 资源。

**关键词:**图像识别;卷积神经网络;迁移学习;贝叶斯优化算法;VGG-16

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2022)0068-04

## Research on CNN Algorithm Based on Bayesian Optimization and Transfer Learning

YU Zhen-tao<sup>1</sup>, LIU Wan-li<sup>2</sup>, YANG Xiao-hui<sup>2</sup>, HUANG Yu-zhen<sup>2</sup>, XU Lei<sup>1</sup>, XIA Ji-an<sup>3</sup>

(1. School of Computer Science and Engineering, Nanjing University of Science and Technology,

Nanjing 210094, China;

2. Nanjing Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing 210014, China;

3. Nanjing Vocational University of Industry Technology, Nanjing 210023, China)

**Abstract:** With the rapid development of deep learning, deep convolutional neural networks have been widely applied in the field of image recognition. However, in practical application, the deeper convolutional neural networks have strong fitting ability and superior performance, but also have the problems of difficult model training and it will take long time. In order to solve this problem, we propose to use Bayesian optimization algorithm to optimize the hyperparameter of convolutional neural network. In the process of continuous iteration, the optimal solution of the hyperparameter vector is obtained, and then the optimal solution parameters are used to train and verify the neural network and generate the optimal classification model. In addition, in order to avoid the drastic adjustment of network weight parameters in each iteration, VGG-16 is used in the bottom layer of the network model architecture for migration learning, which greatly reduces the time of network convergence in one iteration. In this paper, the experiment is based on MNIST and CIFAR-10 datasets, according to the result of experiment, it shows that under the precondition of without sacrificing the accuracy of image classification model, using the Bayesian optimization algorithm only needs a few iterations can search the optimal hyperparameter, greatly saving the time and GPU resources to generate the optimal classification model.

**Key words:** image recognition; convolutional neural network; transfer learning; Bayesian optimization algorithm; VGG-16

## 0 引言

深度学习是一类新兴的多层神经网络学习算法,因其缓解了传统训练算法的局部最小性,引起机器学习

领域的广泛关注<sup>[1]</sup>。近年来随着深度学习的快速发展,卷积神经网络作为一种将人工神经网络和深度学习技术结合的新型人工神经网络方法在图像识别领域

收稿日期:2021-10-14

基金项目:国家自然科学基金(61973161);江苏省科技计划项目(BE2021610);江苏省中医药领军人才资助项目

作者简介:俞圳韬(1996-),男,硕士研究生,研究方向为图像识别;通信作者:刘万里(1972-),男,博士,教授,研究生导师,研究方向为大数据分析、中医内科学脾胃病。

获得了广泛的应用<sup>[2-3]</sup>,尤其对于医学图像的分类有着不错的效果。但是在实际应用中,越深的卷积神经网络<sup>[4]</sup>在拥有强大拟合能力和优越性能的同时,也存在着一个难以避免的问题,就是训练得到一个最优的网络模型往往需要耗费大量的时间和 GPU 资源。

这个问题的出现主要有两方面的原因,一方面,深层次卷积神经网络有着大量的权重参数,更新优化权重参数的计算量十分庞大,使得训练网络本身就需要耗费大量时间;另一方面,分类模型在生成过程中需要对网络的超参数不断迭代优化,并且如果超参数维数较高或者领域范围较大的话,迭代次数更会大大增加,这进一步使得分类模型难以训练。

针对以上问题,该网络模型基于 VGG-16 迁移学习提取图像的特征,使用自定义全连接层进行特征分类,从而大大缩短了训练网络本身所耗费的时间;另外,还使用了贝叶斯优化算法对卷积神经网络的超参数进行优化,该算法大大减少了网络参数迭代的次数,节省了生成最优分类模型耗费的时间和 GPU 资源。

## 1 基于贝叶斯优化和 VGG 特征提取器的网络模型

### 1.1 网络模型架构

常见的卷积神经网络有 AlexNet<sup>[5]</sup>、VGG<sup>[6]</sup>、NIN<sup>[7]</sup>、GoogleNe<sup>[8]</sup>等。AlexNet 使用 ReLU 函数作为激活函数,加快了网络的收敛速度,但是使用了  $7 \times 7$  大卷积核,导致网络参数规模过大而难以训练;VGG 模型通过使用  $3 \times 3$  小卷积核、增加网络规模等方法提升网络的泛化能力,另外,VGG 使用了逐层训练的方法,这使得 VGG 网络模型架构可以很好地提取到图像的低维特征,并抽象成高维特征图。

迁移学习<sup>[9]</sup>可以实现相似任务之间的知识迁移,如果将迁移学习运用在图像识别模型的学习训练中,那么可以在已有的图像识别模型的基础上,使用当前任务对应的图像数据集对该模型继续进行训练,得到一个新的图像识别模型。

基于上述的理论分析,对卷积神经网络模型进行构建。首先,由于 VGG-16 是一个性能优越的特征提取器,因此,提出的网络模型架构使用 VGG-16 的卷积层进行构建,并且去除 VGG-16 的全连接层部分,用来作为模型的底层网络提取图像特征<sup>[10-12]</sup>;其次,提出的网络模型架构使用具有一定宽度和深度全连接层进行构建,用来作为模型的高层分类器,对图像特征进行综合,输出最终的分类结果。其中全连接层的宽度指的是一层全连接层中神经元的个数,全连接层的深度指的是全连接层的层数。提出的网络模型架构具体如图 1 所示。

图中的  $w$  表示全连接层的宽度,  $h$  表示全连接层的深度,VGG16 模型只使用卷积层进行迁移学习,且该模型是使用 imagenet 数据集学习训练得到的。

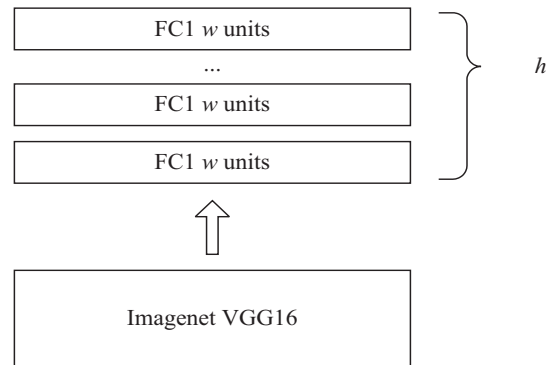


图 1 网络模型架构

### 1.2 贝叶斯优化算法

贝叶斯优化算法<sup>[13-14]</sup>是一种十分有效的全局优化算法,该算法通过设计恰当的概率代理模型和采集函数,只需经过少数几次目标函数评估即可获得理想解。近年来,贝叶斯算法在解决搜索参数最优解问题上有着广泛的应用,赵亚武等人<sup>[15]</sup>使用贝叶斯优化模型优化卷积神经网络的网络结构和超参数,实现对蛋白质二级结构的预测。邓帅<sup>[16]</sup>提出一种基于改进的贝叶斯优化算法的 CNN 超参数优化方法,该方法使用改进的汤普森采样方法作为采集函数,使用改进的马尔可夫链蒙特卡罗算法加速训练高斯代理模型,实现在超参数空间不同的 CNN 框架下进行超参数优化。刘小熊等人<sup>[17]</sup>提出一种基于改进贝叶斯优化算法的故障模式聚类算法,通过结合贝叶斯优化算法中的先验知识来提高算法的可靠性和全局收敛性;将改进的优化算法应用到高维数据最优统计聚类分析中,可快速优化聚类参数,得到全局最优解。

受到上述研究成果的启发,该文提出使用贝叶斯优化算法优化基于 VGG-16 构建的网络模型架构。优化的目标主要为网络的结构和超参数,网络的结构主要体现在对全连接层宽度和深度的优化,网络的超参数主要体现在对学习速率、梯度冲量、正则化系数的优化。

#### 1.2.1 构建优化参数向量

该文指定的网络模型架构底层利用 VGG 网络作为特征提取器,将低维特征提取为高维特征,然后再使用多层全连接对高维特征进一步提取,进行特征分类。因此,指定全连接层的深度和宽度作为要优化的参数;另外,在网络训练过程中涉及到的一些超参数也要进行优化,包括学习速率、梯度冲量和正则化系数。最终生成优化参数向量  $X = [d, w, \alpha, \beta, \lambda]$ ,作为贝叶斯优化算法中的输入,其中  $d$  为全连接层的深度,  $w$  为全连

接层的宽度,  $\alpha$  为学习速率,  $\beta$  为梯度冲量,  $\lambda$  为正则化系数。

### 1.2.2 确定域空间和目标函数

在执行贝叶斯优化算法之前,需要指定采样点所属的参数向量域空间  $D$ ,从域空间  $D$  中采集参数向量作为输入;另外,还需要指定采样点的预测规则,作为采样点  $X_i$  与预测值  $Y_i$  的映射关系,即目标函数。当获取到采样点  $X_i$  后,根据向量中的参数  $d$  和  $w$  确认网络中全连接层的结构;基于向量中的参数  $\alpha, \beta, \lambda$  对网络进行训练,将收敛之后得到的损失函数值作为该采样点的预测值:

$$Y_i = f(X_i) = \text{Loss}(W, X_x) \quad X_i \in D \quad (1)$$

### 1.2.3 贝叶斯优化

该文所述的贝叶斯优化算法首先需要采集  $n_0$  组优化参数,基于优化参数确定卷积神经网络的结构并进行训练,将训练结果作为预测值输出,生成初始候选解集合。然后基于给定的初始候选解集合,使用先验的高斯函数,用来代替未知的待优化的黑盒函数,高斯函数的分布具体由协方差和均值决定:

$$L = \text{GP}(K_x, \mu_x) \quad (2)$$

式中,  $K$  为协方差,  $\mu$  为均值。在每次计算或者更新协方差时,需要指定协方差公式,该文使用高斯核函数作为协方差公式:

$$\begin{cases} k(x_*, x_1) = \alpha_0 \exp(-\frac{1}{2\sigma^2} \|x_* - x_1\|^2) \\ K_x = [k(x_*, x_1), k(x_*, x_2), \dots, k(x_*, x_{n_0})]^T \end{cases} \quad (3)$$

表 1 MNIST 超参数迭代优化情况

迭代次数	深度	宽度	学习率	冲量	正则化系数	Top-1	Top-5
1	2	1 024	0.001 2	0.962 3	$2.15 e^{-10}$	90.32	91.33
2	2	2 048	0.001 6	0.952 8	$3.22 e^{-5}$	91.97	92.12
3	3	1 024	0.003 4	0.891 2	$4.22 e^{-9}$	92.33	94.32
4	3	2 048	0.002 3	0.871 7	$2.72 e^{-8}$	98.71	99.10
5	3	3 096	0.004 1	0.976 1	$5.19 e^{-7}$	98.42	99.02

表 2 CIFAR-10 超参数迭代优化情况

迭代次数	深度	宽度	学习率	冲量	正则化系数	Top-1	Top-5
1	1	1 024	0.002 1	0.957 3	$4.99 e^{-12}$	88.15	90.52
2	2	1 025	0.001 3	0.961 8	$6.19 e^{-8}$	89.32	91.43
3	2	2 048	0.002 1	0.781 2	$5.63 e^{-11}$	90.92	92.37
4	3	3 096	0.001 9	0.811 7	$1.72 e^{-9}$	92.88	94.72
5	3	3 096	0.003 1	0.964 2	$3.14 e^{-8}$	92.97	94.38

基于上述结果可知,在数据集 MNIST 上进行实验

式中,  $k$  为协方差公式,  $x_*$  为优化的参数。然后从域空间  $D$  中采取一组参数向量并根据目标函数预测输出值,则当前采样数据为:

$$T = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\} \quad (4)$$

式中,  $(X_n, Y_n)$  为新采样添加的一组候选解,然后基于新的采样数据更新高斯函数的均值和协方差<sup>[18-19]</sup>,即对黑盒函数<sup>[20]</sup>进行优化,从而得到该函数的概率分布  $p(f(X_i) | T)$ ,根据更新后均值和方差可以计算采集函数  $\mu(x)$ ,计算采集函数的极大值确定下一个采样点:

$$X_{n+1} = \text{argmax}_x(\mu(x)) \quad (5)$$

然后再次添加一组新的候选解  $(X_{n+1}, Y_{n+1})$ ,再次更新高斯函数的均值和方差。上述步骤不断地迭代,高斯替代函数不断地被优化,迭代到一定次数后,输出优化后高斯函数的极大值对应的输入  $X_p$ ,即最优的参数向量。

## 2 实验与结果分析

首先,分别在数据集 MNIST、CIFAR-10 上使用贝叶斯优化算法迭代优化卷积神经网络的超参数,超参数包括全连接层深度、全连接层宽度、学习速率、梯度冲量以及正则化系数,并以分类模型的准确率作为最终评价指标。超参数的迭代优化结果如表 1 和表 2 所示。

时,前 3 次迭代模型的准确率都有所提高,直到第 4 次

迭代准确率相对饱和而收敛;在数据集 CIFAR-10 上进行实验时,前 4 次迭代模型的准确率都有所提高,直到第 5 次迭代准确率相对饱和而收敛,由此,可以得出结论,使用贝叶斯优化算法只需要几次迭代就可以找到最优的超参数向量。另外,前几次迭代的准确率较低,主要原因是超参数没有调整至最佳,而在模型收敛之后,数据集 MNIST 上的准确率小于 99%,数据集 CIFAR-10 上的准确率小于 93%,主要是因为受到网络模型本身泛化能力的限制。

其次,在数据集 MNIST 上,分别使用贝叶斯优化算法、随机搜索方法、网格搜索方法,多次迭代搜索卷积神经网络的最优超参数,搜索结果对比如图 2 所示。

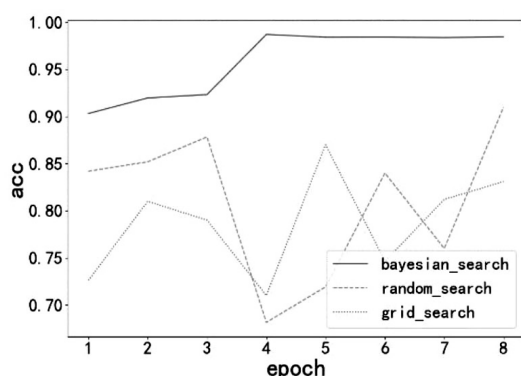


图 2 贝叶斯/随机/网格搜索结果对比

基于上述结果可知,相较于随机搜索和网格搜索,贝叶斯优化算法在搜索最优超参数的过程中,每一次迭代都可以有效提升模型的分类准确率,搜索显得更有方向性,可以实现收敛得到超参数最优解。

### 3 结束语

深层卷积神经网络在医学图像识别领域有着广泛的应用和研究前景,该文基于 VGG-16 的网络模型架构进行迁移学习,并且融合贝叶斯优化算法对卷积神经网络的超参数进行优化,极大地缩短了生成最优分类模型耗费的时间。在未来的工作中,将会结合舌象图像构建中医体质分类器,并且将会对网络模型架构进一步优化。

#### 参考文献:

[1] 孙志军,薛磊,许阳明,等.深度学习研究综述[J].计算机应用研究,2012,29(8):2806-2810.  
[2] 许可.卷积神经网络在图像识别上的应用的研究[D].杭州:浙江大学,2012.

[3] 尹宝才,王文通,王立春.深度学习研究综述[J].北京工业大学学报,2015,41(1):48-59.  
[4] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.  
[5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.  
[6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409: 1556, 2014.  
[7] LIN M, CHEN Q, YAN S. Network in network[J]. arXiv: 1312:4400, 2013.  
[8] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston, MA: IEEE, 2015: 1-9.  
[9] 胡满满,陈旭,孙毓忠,等.基于动态采样和迁移学习的疾病预测模型[J].计算机学报,2019,42(10):2339-2354.  
[10] 宋光慧.基于迁移学习与深度卷积特征的图像标注方法研究[D].杭州:浙江大学,2017.  
[11] 许景辉,邵明烨,王一琛,等.基于迁移学习的卷积神经网络玉米病害图像识别[J].农业机械学报,2020,51(2):230-236.  
[12] 张建华,孔繁涛,吴建寨,等.基于改进 VGG 卷积神经网络的棉花病害识别模型[J].中国农业大学学报,2018,23(11):161-171.  
[13] HUANG C, REN Y, MCGUINNESS E K, et al. Bayesian optimization of functional output in inverse problems[J]. Optimization and Engineering, 2021, 22: 2553-2574.  
[14] 崔佳旭,杨博.贝叶斯优化方法和应用综述[J].软件学报,2018,29(10):3068-3090.  
[15] 赵亚武,刘毅慧.基于优化的卷积神经网络的蛋白质二级结构预测[J].计算机应用与软件,2021,38(7):147-152.  
[16] 邓帅.基于改进贝叶斯优化算法的 CNN 超参数优化方法[J].计算机应用研究,2019,36(7):1984-1987.  
[17] 刘小雄,武燕,史静平,等.一种故障诊断的贝叶斯优化算法研究[J].计算机应用研究,2009,26(1):129-131.  
[18] 何志昆,刘光斌,赵曦晶,等.高斯过程回归方法综述[J].控制与决策,2013,28(8):1121-1129.  
[19] 孙斌,姚海涛,刘婷.基于高斯过程回归的短期风速预测[J].中国电机工程学报,2012,32(29):104-109.  
[20] TENNE Y. Machine-learning in optimization of expensive black-box functions[J]. International Journal of Applied Mathematics and Computer Science, 2017, 27(1): 105-118.